2004-03

Speech recognition software: an alternative to reduce ship control manning

Kuffel, Robert F.

Monterey California. Naval Postgraduate School

http://hdl.handle.net/10945/1662

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

SPEECH RECOGNITION SOFTWARE:
AN ALTERNATIVE TO REDUCE SHIP CONTROL MANNING

by

Robert F. Kuffel

March 2004

Thesis Co-Advisors: Russell Gottfried
Monique P. Fargues

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2004	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Speech Recognition Software; An Alternative to Reduce Ship Control Manning			**5. FUNDING NUMBERS**
6. AUTHOR(S) Robert Kuffel, LT USN			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES: The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This study identifies factors affecting the performance of commercial-off-the-shelf speech recognition software (SRS) when used for ship control purposes. After a review of research in the feasibility and acceptability of SRS-based ship control, the paper examines the effects of:

- A restricted vocabulary versus a large vocabulary,
- Low experience level conning officers versus high experience level conning officers,
- Male versus female voices,
- Pre-test training on specific words versus no pre-test training.

Controlled experimentation finds that:

- The experience level of a conning officer has no significant impact on SRS performance.
- Female participants experienced more SRS errors than did their male counterparts. However, in this experiment, only a limited number of trials were available to assess a difference.
- SRS with restricted vocabulary performs no better than SRS with large vocabularies.
- Using the software "correct as you go" feature may impact software performance. Following the user profile establishment, individual user training on two specific words reduces error rates significantly.

This study concludes that SRS is a viable technology for ship control and merits further testing and evaluation.

14. SUBJECT TERMS: Speech Recognition, Voice Recognition, Ship Maneuvers, Standard Commands, Ship Handling, Commercial-Off-The-Shelf Software, Voice Activated Control System, Manpower Reducing Technology			**15. NUMBER OF PAGES** 62
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

SPEECH RECOGNITION SOFTWARE; AN ALTERNATIVE TO REDUCE SHIP CONTROL MANNING

Robert F. Kuffel
Lieutenant, United States Navy
B.A., California State University, 1996

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS AND OPERATIONS

from the

NAVAL POSTGRADUATE SCHOOL
March 2004

Author: Robert F. Kuffel

Approved by: Russell Gottfried
Thesis Co-Advisor

Monique Fargues
Thesis Co-Advisor

Dan C. Boger
Chairman, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This study identifies factors affecting the performance of commercial-off-the-shelf speech recognition software (SRS) when used for ship control purposes. After a review of research in the feasibility and acceptability of SRS-based ship control, the paper examines the effects of:

- A restricted vocabulary versus a large vocabulary,
- Low experience level conning officers versus high experience level conning officers,
- Male versus female voices,
- Pre-test training on specific words versus no pre-test training.

Controlled experimentation finds that:

- The experience level of a conning officer has no significant impact on SRS performance.
- Female participants experienced more SRS errors than did their male counterparts. However, in this experiment, only a limited number of trials were available to assess a difference.
- SRS with restricted vocabulary performs no better than SRS with large vocabularies.
- Using the software "correct as you go" feature may impact software performance. Following the user profile establishment, individual user training on two specific words reduces error rates significantly.

This study concludes that SRS is a viable technology for ship control and merits further testing and evaluation.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

I. BACKGROUND

A. INTRODUCTION

In recent years, the U.S. Navy has begun to search for ways to decrease the number of personnel necessary to operate a ship at sea. Initiatives such as “Smart Ship” design and the ongoing “Optimal Manning trials” are designed to show that ships can operate at sea with reduced manning. [Ref. 1] Ship designers actively seek-out manpower-saving, technology-based options. Technological advances have not only reduced manning, but in many cases enhanced the ability of watchstanders to conduct their duties. For example, engineering watchstanders use handheld computers to record engineering plant data that is then downloaded to a computer which automatically generates required reports. Combat watchstanders employ touch-screen technology and automated display screens to speed the process of data entry and display. More efficient and accurate computerized navigation systems enable quartermasters to plan and plot ship movements. [Ref. 2]

Ship control, however, remains an area that seems relatively untouched by technological advances, as traditions developed long before the birth of the U.S. Navy still remains in place. This thesis explores one technology alternative, building upon previous research in investigating the viability of using speech recognition software (SRS) aboard naval vessels for ship control purposes, and analyzing the system’s potential in eliminating the need for two bridge watchstanders, the helmsman and the lee helmsman.

Chapter I establishes a foundation of knowledge by examining the background and historical information related to speech recognition technology. Chapter II describes how SRS is applicable to naval vessels and considers potential barriers to its employment. Chapter III delineates the methodology utilized in an experiment designed to show sources of performance variation and potential avenues to reduce SRS error rates. Experimental results and their analysis are presented in Chapter IV; and finally, Chapter V summarizes findings, makes recommendations and proposes areas of future research.

B. SPEECH RECOGNITION TECHNOLOGY

Speech Recognition Software (SRS), also called Voice Recognition Software (VRS), enables a computer to convert a spoken word (an acoustic signal) into text which is represented within the computer by binary digits. At the heart of the software is an analog-to-digital converter which digitizes the incoming analog signal and divides it into 10 to 20 millisecond frames. [Ref. 3] These frames are then compared to a digital library stored in memory.

> Speech recognition systems focus on words and the sounds that distinguish one word from another in a language. Those sounds are called phonemes. The words "seat," "beat," and "Cheat" are different words because, in each case, the initial sound is recognized as a separate phoneme in English. [Ref. 4]

The lexicon library contains phoneme models which define the pronunciation of a word as well as its length. It may also contain multiple pronunciations of the same word to account for regional differences in pronunciation. The "matching" process does not seek out an exact phoneme match but rather looks for the best match. Using a procedure known as Stochastic Processing, incoming signals are compared to a set of potential candidates using Hidden Markov Models (HMM), which provide a way to represent the likelihood of a transition from one phoneme to the next in a given word.

> These comparisons produce a probability score indicating the likelihood that a particular stored HMM reference model is the best match for the input. [Ref. 5]

This approach allows the computer to focus on the shape of the vocal tract and make allowances for extraneous information and slight differences that occur each time a given word is spoken.

The adaptability of SRS technology is one of its strengths. SRS technology may be incorporated into a Voice Activated Command System (VACS) which uses the digital signal output to control other electronics or machinery. SRS software also has several parameters that can be adjusted based on the needs of the user. These parameters and the range of the adjustment are shown in Table 1. The parameter settings utilized in this thesis research include speaking mode and style, user enrollment, vocabulary and language sensitivity.

PARAMETER	RANGE
Speaking Mode	Isolated to Continuous
Speaking Style	Scripted to Spontaneous
Enrollment	Speaker Dependent to Speaker Independent
Vocabulary	Small to Large
Language model	Finite State to Context Sensitive

Table 1. SRS Parameters from Ref. 6

This thesis focuses on the continuous speaking mode, allowing the user to speak naturally as opposed to pausing between each word when using an isolated speaking mode. All verbal orders given on the bridge of a ship consist of short phrases that are spoken naturally. For this reason, a system using an isolated speaking mode would be ineffective for ship control purposes.

A scripted speaking style is designed for users who will read information and avoid verbal irregularities such as verbal pauses ("uhs" and "ums"). A spontaneous speaking style is more characteristic of the bridge of a ship and hence will be explored in this thesis. The software can be "trained" to filter out verbal irregularities as described below.

SRS software is available in Speaker-independent and Speaker-dependent varieties. This thesis focuses on a Speaker-dependent system. A speaker independent system is capable of recognizing the voices of many different speakers, whereas a speaker dependent system is trained to specific voices. [Ref. 7] The process of training the system to a specific individual is often referred to as "setting up a user profile". Each user sets up a profile by repeating a set of words or phrases multiple times so that the software can create a baseline model of the user's speech patterns. The model allows for a certain degree of variability such as pitch and or pace changes, raspy voices, and other non standard speech tendencies and it accounts for slight differences that may occur each time the speaker speaks.

The size of the vocabulary utilized by most SRS is adjustable. The vocabulary, sometimes referred to as the library, is a list of words that the software can recognize. Small vocabularies contain less than 1000 words while most large vocabularies can handle up to 70,000 words. The size of the vocabulary selected is dependent on the task to be accomplished. Dictating a legal memo for example would require a much larger vocabulary than generating a basic grocery list. SRS is more efficient and accurate when a small vocabulary is used because there are fewer alternatives from which the computer chooses. [Ref. 8] For this reason, this study examines SRS performance using a small vocabulary. The specialized orders used for driving naval ships are called Standard Commands. They consist of a very limited number of words in a specific order (See Appendix A) which make them ideally suited for use in a limited vocabulary SRS. Chapter II will provide more details regarding the use of Standard Commands.

Another parameter difference that can exist between speech recognition systems is the use or non-use of a language model. A context-sensitive language model will inspect the surrounding words in order to determine which word to insert. Often a statistical language model determines the estimated frequency of word usage and selects the most probable sequence of words. [Ref. 9] The SRS software used in this study includes a built-in language model.

C. PRESENT DAY SPEECH RECOGNITION SOFTWARE USES

Advancements in speech recognition technology have made it useful for a variety of commercial and private uses including: dictation, personal computer interfaces, inventory maintenance, automated telephone services and special purpose industrial applications. [Ref. 10] Even items that are as small as cellular phones and personal data assistants are now capable of recognizing hundreds of words. In the home, speech recognition software simplifies the man-machine interface by allowing for verbal control of such items as televisions, household lighting, environmental controls, [Ref. 11] and stereo systems. [Ref. 12]

The Department of Defense has also taken an interest in the applications of SRS technology. SRS hands-free, heads-up nature makes it ideal for military applications. In

addition, its resulting man-power saving attributes have led to its use in training simulators. For example, the U.S. Navy Surface Warfare Officers School (SWOS) in Newport, Rhode Island, now uses a Voice Activated Bridge Simulator to teach ship handling skills to newly commissioned officers. In Groton, Connecticut, the U.S. Navy Submarine School has reduced staffing needs for its Virtual Submarine trainer by introducing SRS technology. Personnel involved with submarine simulator operations perceive value in SRS.

> Voice recognition and synthesis software allow the student to interact with a computer-generated navigator, helmsman, and engineering officer of the watch. The students can issue commands that the computer sub recognizes and responds to just as humans would. [Ref. 13]

Note that these systems are speaker independent and do not require users to set up profiles before use. As a result, they may be more susceptible to errors caused by accents and rises in pitch due to excitement in response to simulated hazardous situations. This reinforces a standard, consistent form among conning officers.

> In the training environment, there is an added value to using speaker independent systems. They force students to learn to remain calm on the bridge and give verbal orders in a clear, crisp voice. [Ref. 14]

However, natural variability in human performance is a reality in the fleet. A robust, reliable VACS will have to respond to orders accurately. To do so it will need to account for speaker dependent variation.

D. PREVIOUS SRS RESEARCH

In February 2001, Ingall’s Shipbuilding conducted an experiment to test the usefulness of an Integrated Bridge System (IBS) that they had developed. An integral part of the IBS was VACS. Even though the purpose of their test did not focus on VACS, the study yielded insight regarding SRS.

- Participants in the study preferred VACS to normal control methods but agreed that there needed to be the ability for the conning officer to take manual control if necessary.
- The testing also revealed a need for a standard command vocabulary to be built into the VACS.

- Finally, tests showed that there needed to be some type of resolution for a misinterpreted command so that the system would not take incorrect action or fail to respond. [Ref. 15]

A June 2003 Naval Postgraduate School (NPS) study developed an experiment designed to show the reliability of commercial-off-the-shelf (COTS) speech recognition software [Ref. 16]. It used a commercially available SRS called Dragon Naturally Speak, version 6.0 (DNSv6.0) to record the verbal orders of conning officers who were driving a simulated ship. DNSv6.0 is a continuous, spontaneous, speaker-dependent system that utilized a large vocabulary of over 20,000 words and had a built in language model.

The experiment took place in the Marine Safety International (MSI) San Diego, California, shipboard simulator and used experienced ship handlers as test subjects. Using a wireless microphone, test subjects transmitted verbal commands to a nearby laptop computer which then used DNSv6.0 to convert the verbal orders to text. These text files were later analyzed for errors made by the software [Ref. 17]. The research conclusions provide insight into the use of SRS for ship control purposes.

Results varied based on who used the SRS. Some subjects seemed to be able to speak more clearly than others and therefore had fewer errors. The study hypothesized that additional system training for each test subject could potentially eliminate some of this variability.

Second, the results demonstrated that the operational scenario had no impact on the system performance. In other words, it did not matter if the test was conducted on a simulated Destroyer, Frigate or Cruiser. Further, it did not matter if the simulated ship was entering port, leaving port, engaged in open ocean transit or any combination of these.

This study also revealed that the ambient noise level of the setting influenced SRS performance. While the SRS profiles were developed in a relatively quiet room, the experiment was conducted in a simulator with increased ambient noise. Subsequent analysis pointed out that initial profile development and experiment conduct for each test subject should take place in the same setting to "teach" the system to filter out any ambient noise present. [Ref. 18]

Based on the lessons learned from Ingall's Ship Building and previous NPS research, follow-on research is necessary to better determine the viability of COTS SRS as used for ship control purposes. In the chapters that follow, this thesis documents that follow-on research. First, however, it is important to discuss why this research is of interest to the U.S. Navy.

THIS PAGE INTENTIONALLY LEFT BLANK

II. APPLICABILITY TO NAVAL VESSELS

A. WATCHSTANDING

A U.S. Naval ship typically has eight watchstanders manning the bridge while underway at sea. The watchstander positions and their duties can be found in Table 2 [Ref. 19]. Some ships may modify this list by adding to or subtracting from these positions based on the vessel traffic density, visibility and other navigationally significant circumstances.

POSITION	**DUTIES**
Officer of the Deck (OOD)	Represents the Captain and makes decisions regarding the safe operation of the ship.
Junior Officer of the Deck (JOOD)	OOD in training - usually handles tactical communications and computes maneuvering solutions.
Conning Officer (CONN)	Issues rudder and propulsion orders to the helmsman and lee helmsman.
Boatswains Mate of the Watch (BMOW)	Supervises the enlisted watch team. Usually a qualified master helmsman.
Quartermaster of the Watch (QMOW)	Navigates the ship and keeps the deck log. Usually qualified as a helmsman.
Helmsman	Carries out the rudder orders of the conning officer by steering the ship.
Lee Helmsman	Carries out the propulsion orders of the conning officer by making speed adjustments.
Phone Talker	Maintains communications between vital stations.

Table 2. Bridge Watch Stations

This thesis will focus on the Conning Officer, the Helmsman, and the Lee Helmsman watchstanders. The interaction among these three individuals is meant to ensure that no order is misunderstood. Each order given by the conning officer is repeated back verbatim to ensure complete understanding by the helmsman or lee helmsman. In this fashion, immediate corrective action can be taken if any order is

misunderstood. This system of repeat-backs also serves two other important purposes. First, it aids accountability by enabling the quartermaster of the watch to record each of the conning officer's orders in the deck log. Second, it helps everyone on the bridge watch team maintain awareness regarding the status of ship maneuvers.

Orders issued by the conning officer are standard commands, the exact words and the sequence of which are formalized on all naval warships. The list of standard commands can be found in Appendix D. Note that it is a relatively small vocabulary totaling fewer than 100 words. The exact number depends on the ship type. This small vocabulary makes ship driving a strong candidate for speech recognition software implementation. Newly commissioned Surface Warfare Officers (ship drivers) undergo extensive training to learn to use the standard commands properly. By the time an officer has completed the qualification process, the standard commands are as second-nature as speaking.

B. REDUCED MANNING ISSUES

In pursuit of reducing the manpower requirements to operate a ship at sea, the Navy also reduces ship life-cycle costs [Ref. 20]. There are however, additional reasons for reducing ship manning requirements. Many ships in the Navy today are unable to meet their allocated manning levels and watch station requirements. [Ref. 21] An undermanned ship is more prone to manpower fatigue, has little room for training replacement personnel and has the risks associated with reduced redundancy, potentially affecting the safety of the ship itself. As of 2001, ninety-one percent of all mishaps reported to the Naval Safety Center were caused by human error. In many of these cases, improper training or fatigue played a role. [Ref. 22] In addition to saving money, reducing manning requirements through the installation of technology may also alleviate current shortages, thereby making ships safer.

The course of action prescribed by the Naval Transformation Roadmap is to "…insert technology to carry out operations in ways that profoundly improve current capabilities and develop desired future capabilities." [Ref. 23] Aligned with this guidance is the Smart Ship program which was developed to reduce shipboard personnel numbers by inserting technology that replaces watchstanders. The results of this initiative have

been so successful that newer ships are being designed with more technology and smaller crew sizes. A prime example is the Navy's Littoral Combat Ship (LCS) which is still being developed, but project decision makers envision dramatically reduced manning levels. Senior naval officials acknowledge that it is only a matter of time before this move to replace watchstanders with technology affects the way navy warships man their bridge watch teams through "… a significant reduction in bridge manning needs." [Ref. 24]

In his 2004 guidance the Chief of Naval Operations stated, "As our Navy becomes more high tech, our workforce will get smaller and smarter." [Ref. 25] His words rang true in January 2004 as 1,900 billets were trimmed from the fleet. The 2005 budget includes further plans to eliminate sailor and officer jobs throughout the Navy. [Ref. 26] As Admiral Clark puts it, "…we do not want to spend one extra penny for manpower we do not need." [Ref. 27] The cuts are enabled by the elimination of redundant functions and the installation of manpower-saving technology. The CNO wants to "look at options for carrying out midterm modernization on all the Navy's surface ships." [Ref. 26]

C. TECHNICAL FEASABILITY & IMPLEMENTATION

Use of SRS for ship control purposes could eliminate the helmsman and lee helmsman watch stations during open-ocean steaming. The purpose of this study is to assess the software technology's ability to replace these watchstanders. VACS could be faster and more accurate than a human watchstander as well.

> … Voice Recognition system devices (the system's hardware) would be physically installed into the ship's current Ship's Control Console… connected electronically from the SCC to the engineering propulsion and steering systems for immediate responses to the Conning Officer's orders. The Conning Officer and Officer of the Deck would both be equipped with cordless microphone headsets that would have attached activation switches allowing navigational commands to be given on demand. [Ref. 28]

In order to maintain the current checks and balances between the Conning Officer and the helmsman or lee helmsman,

> …the VR system would be equipped with a series of speakers installed throughout the ship's bridge. The purpose of the bridge speakers is to broadcast orders given by the Conning Officer as well as the repeat-back by the VR system. This enables all bridge watch standers to hear the orders and repeat-backs, allowing them to maintain situational awareness as to how the ship is being driven and to anticipate the ship's actual movements. The speakers will also serve to provide a means for the VR system to repeat back the ordered command. [Ref. 29]

The system could also be programmed to ask the conning officer to repeat the command (e.g., "Orders to the Helm?") if the system did not find a match in its standard command library.

A final issue to consider when implementing a VACS is casualty control. As suggested by the Ingall's IBS testing a quick disconnect button is necessary so that any time the need arises the ship can return to manual mode. As with most other vital shipboard equipment, a monitoring and alarm panel would enable instant fault detection, prompting bypass of VACS. Upon bypass of VACS, another bridge watchstander could step in and execute the functions of helmsman and or lee helmsman.

Implementation of SRS on the bridge of Navy ships is technically feasible and may actually prove more efficient than the manual control methods currently in place. Further, such a system causes very few procedural changes to bridge watch standing while aiding the ongoing effort to reduce the number of personnel required to operate a ship at sea. There is however, resistance to the idea of using SRS onboard naval ships. This resistance is well documented.

D. PSYCHOLOGICAL BARRIERS

One of the greatest obstacles to implementing this new technology is the human resistance to change. [Ref. 30] The Navy is an organization based on longstanding traditions with bureaucratic forces that encourage maintaining the status quo. Leaders that fail to uphold the traditional way of doing things are seen as "risk takers". However, it is precisely these "risk takers" who may enable innovations and progress in the Navy. [Ref. 31]

In an October 2002 study, 110 Surface Warfare Officers ranging in rank from Ensign to Captain were asked if they would allow a voice activated control system aboard ships. Eighty percent of these said that they would allow it. Many added that initially its use should be limited to certain circumstances. [Ref.32] The condition most often stated as a qualifier to VACS use was that the ship be in open ocean transit with no other vessels nearby. Most respondents also said that with time and proven reliability restrictions to use could be relaxed. [Ref. 33]

The remaining twenty percent of respondents stated that they would not endorse the use of VACS onboard Navy ships. [Ref. 34] Reasons given for not wanting to implement VACS included the perceived increased risks associated with "letting a computer drive the ship" and the lack of human interaction between the helmsman and conning officer. [Ref. 35] Respondents suggested that having a helmsman in the loop added an additional safety check in driving the ship, because a good helmsman may catch an error made by the conning officer.

The first of these two arguments has little merit as computers are used for a number of risk inherent activities. The Aegis computer system can be trusted to defend the ship in battle. Analogously, a VACS with similar redundancies and safeguards could relay the conning officer's orders to the engines and rudders. The second argument regarding helmsman and conning officer interaction has some validity. However, even if the helmsman were not present, other personnel on the bridge could alert the conning officer to an erroneous decision; specifically, the Officer of the Deck or an alert Quartermaster of the watch. Additionally, the current speech-to-text capability of SRS will alleviate quartermaster deck log duties, allowing for greater oversight.

A conning officer may not understand how a VACS system works and therefore feel less control over it than a human helmsman. By the virtue of their positions, Naval Officers are used to being in control and the idea of relinquishing some of that control may be unnerving. With exposure to the system over time and proven reliability, VACS use can overcome the psychological barriers that reside in some Naval Officers.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

A. EXPERIMENT OBJECTIVE

The objective of this study is to identify factors which affect the performance of specific commercial-off-the-shelf speech recognition software when used for ship control purposes. Specific factors examined include the effects of:

- A restricted vocabulary versus a large vocabulary,
- Low experience level conning officers versus high experience level conning officers,
- Male versus female voices,
- Pre-test training on specific words versus no pre-test training.

The study builds upon previous SRS research and uses data from a prior experiment to examine the relevance of the above factors.

B. EXPERIMENTAL SETTING

1. Prior Research

As outlined in Chapter II, prior experimentation with SRS sought to determine factors that affected error rates. The necessity for this follow-on study is grounded in the need to build upon that previous experimentation.

- An SRS with the default 20,000 word vocabulary, utilized in the previous experimentation, may not have been well matched to the conning application under consideration, due to its limited vocabulary requirement. This study analyzes the impact of replacing the large vocabulary with a small restricted conning vocabulary.
- Test subjects in the previous study were all very proficient male ship handlers each with over ten years of ship driving experience. Actual ship drivers in the fleet are usually newer officers, of both genders, with only limited experience. The higher experience level of the previous test subjects or their gender may have biased the resultant data. The current study uses test subjects with both high levels and low levels of ship driving experience to determine what impact experience level has upon the SRS performance. In addition, female test subjects are introduced, although specific SRS performance variation did not drive experiment design.
- The prior SRS study included no additional system training after the establishment of each test subject's profile. SRS manufacturers claim that

additional system training will improve the accuracy of the SRS [Ref. 36]. An absence of additional training may cause an increased number of errors. This current study addresses the issue by incorporating pre-experiment system training for some test subjects to determine its value in making SRS more accurate.

- Profile establishment in the earlier study took place in the control room while actual testing was conducted in not only the control room but the simulator as well. It could be argued that the profile established in the control room was less effective in the simulator because the ambient noise levels and acoustic qualities varied between these two locations. Ambient noise measurements revealed a 16 dB difference between the two rooms. [Ref. 37] Further, an argument could be made that the control room, does not accurately reflect the actual noise levels experienced on the bridge of a navy ship. It lacks the many electronic navigation devices present on the bridge of a navy ship and in the simulator. The current study conducts all profile establishment and testing in the simulator.

These issues and their implications for individual influence and or combined interaction justify a re-examination of the sources of error to COTS SRS and form the basis for this study.

2. Current Research

While there are differences between this study and the previous investigation, it is also important to discuss the similarities. For example, it was important to hold constant in this study many of the details of the previous one so that a valid statistical comparison between the two can be made. The main difference between the two studies is the size of the SRS vocabulary. Other factors including the COTS SRS software, the experimental setting, the basic test procedure, and the equipment resembled the previous work as closely as possible. Just as in the prior SRS study, the experiment was conducted with the support of Marine Safety International (MSI) facilities using Dragon Naturally Speaking Version 6.0 (DNSV6.0).

3. Marine Safety International

Marine Safety International provides ship handling, Bridge Resource Management (BRM), Electronic Chart Display Information System (ECDIS), Integrated Bridge Systems (IBS) and Automatic Radar Plotting Aids (ARPA) training courses for

the U.S. Navy, the U.S. Coast Guard, MSC, NOAA, International Navies and coastal patrols. There are three MSI locations within the United States; San Diego, CA; Norfolk, VA; and Newport, RI. The equipment at each location is identical and all training is based upon a common curriculum. Each facility features both a bridge wing simulator and a full mission bridge simulator. Only the bridge wing MSI simulator in Newport was utilized for this study. Figure 1 shows the floor plan of MSI Newport.

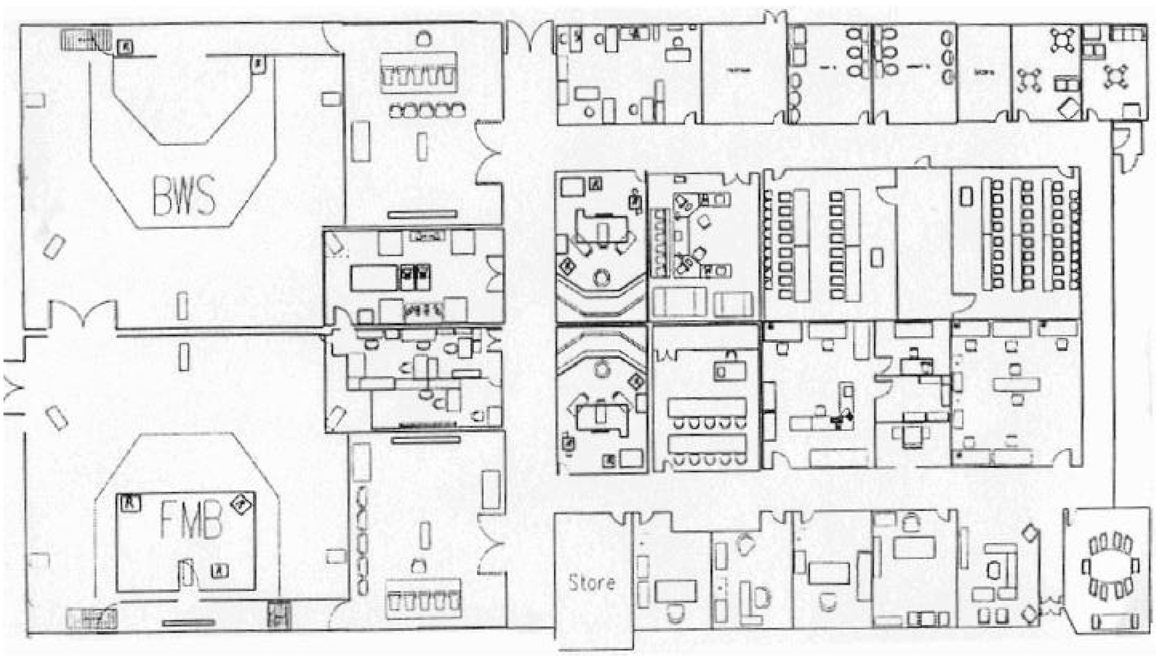

Figure 1. MSI Newport, RI, Floor Plan from Ref. 38

Note that the previous experiment took place at the San Diego MSI and not in Rhode Island. [Ref. 39] However, as stated above, all of the MSI facilities are sufficiently similar with the only detectable difference being the layout of the building's floor plan. [Ref. 38] Even the amount of ambient noise present in the simulator at both locations is comparable. Measurements taken with a Type 2 dB-A sound level meter revealed an ambient noise level of 64.8 dB in the previous study [Ref. 40] while the ambient noise measurements were 66.2 dB in the Rhode Island bridge wing simulator. [Ref.41] This slight difference is acceptable and very realistic when put into the context

of actual ship driving in which ambient noise levels will vary significantly. Additional background noise such as rain, fog horns, etc. can be added to the simulation but this feature was not used during either experiment. All data collection and test subject profile establishment for this current study took place in the simulator with the baseline 64.8 dB sound level. [Ref. 42]

4. Dragon Naturally Speaking Version 6.0

This study uses the exact same speech recognition software as the previous experiment, Dragon Naturally Speaking Version 6.0 (DNSV6.0) which has the following characteristics:

- Continuous Speech Recognition capabilities,
- Speaker dependence,
- Variable vocabulary that allows the user to select the size of the vocabulary desired or to create a specialized vocabulary,
- Spontaneous speech capabilities,
- User-friendly graphic interfaces to facilitate profile set up and application use.

This software is designed to achieve a 90 to 98 percent accuracy rate for most users according to its manufacturer. DNSV6.0 has been top ranked seven times by SRS Software reviewers and this current version is recognized to be superior. [Ref. 42]

5. Test Subjects

Table 3 contains the test-subject data, featuring ten test subjects, five from the MSI staff and five from the Naval Surface Warfare Officer's School (SWOS). MSI test subjects were all retired Navy Captains each with over fifteen years ship handling experience and a surface warfare qualification. Test subjects from SWOS were pre-department head level surface warfare qualified lieutenants each with fewer than four years of ship handling experience. Two of the low experience level test subjects were female. All other test subjects were male.

SUBJECT	GENDER	EXP LVL	Source	SWO
I	M	Low	SWOS	Yes
II	M	Low	SWOS	Yes
III	M	High	MSI Staff	Yes
IV	M	High	MSI Staff	Yes
V	M	High	MSI Staff	Yes
VI	M	High	MSI Staff	Yes
VII	M	Low	SWOS	Yes
VIII	M	High	MSI Staff	Yes
IX	F	Low	SWOS	Yes
X	F	Low	SWOS	Yes

Table 3. Test Subject Data

6. Experimental Procedure

Test subjects were randomly scheduled in two hour blocks, as shown in the MSI/NPS Test document included in Appendix B. Each test subject received the brief included in Appendix C upon arrival at MSI. Following the brief, test subjects moved into the simulator where they established DNSV6.0 speech profiles. In addition to the standard profile establishment, all but three test subjects underwent specialized training on two words that the experiment revealed as having a high incidence of error, “rudder” and “starboard”. Further details are given in Chapter IV.

Next, as conning officers of a CG-47 class Guided Missile Cruiser, the test subjects participated in three different scenarios. To ensure no bias based on the scenario order, the sequence in which these scenarios were presented was randomized and varied for each test subject. Each scenario included approaching a pier and then getting underway from that same pier. Test Subjects wore a SHURE ULX/S Standard Wireless Microphone and issued all orders verbally. The wireless microphone transmitted the verbal commands to a Sony VAIO FX250 Laptop computer located in the control room, approximately 20 feet away. The ULX/S has an RF carrier Frequency Range of 554 to 865 MHz with an effective range of 100 meters. The VAIO laptop was loaded with DNSV6.0 and converted all verbal orders into text for analysis.

During the simulator trials, test subjects received no repeat-backs of orders. Although this is a distinct departure from actual conning procedures, use of repeat backs and acknowledgements yield no insight into SRS performance. A system operator in the control room performed helm and lee helm functions.

C. EXPERIMENT EXPECTATIONS

- **E1: SRS with smaller vocabularies are more accurate (fewer errors) than SRS with large vocabularies**

A major experimental expectation of this study to was whether a small SRS vocabulary produces fewer errors than a large SRS vocabulary. As discussed, the smaller vocabulary results in fewer choices for the software when attempting to match spoken words to the library. This in turn should result in fewer software errors due to misinterpretation.

- **E2: The experience level and/or gender of the SRS user will have no impact on the performance of SRS**

For acceptable operability in the fleet, the experience level of the user should have no impact on the SRS performance. As long as conning officers use standard commands, the system should be able to recognize the verbal orders and convert them to text regardless of the conning officer's age, gender, or experience level. One exception to this may be caused by stress-related pitch elevations in the user's voice. Less-experienced conning officers may tend to be more easily excited while ship driving. To counter this effect during testing, the ship driving scenarios have a very low degree of difficulty and each test subject is instructed to remain calm throughout the scenario as their ship driving abilities are not the focus.

- **E3: Test subjects who undergo additional SRS training will have a lower error rate than those who do not undergo the additional training**

Both ScanSoft, Inc. [Ref. 43] and conclusions from the previous SRS study suggest that additional training prior to SRS use improves accuracy. It is therefore expected that test subjects who receive additional training will experience a lower error rate than those that do not perform the extra training.

Experimentation was conducted over a three day period beginning October 27, 2003. No problems with hardware or software were encountered during the test procedure, and all data was successfully compiled for analysis in the next chapter.

IV. ANALYSIS

A. DATA RESULTS

A total of 30 trials were conducted (10 test subjects with 3 trials each). Appendix D contains the full data worksheet. Analysis suggests four types of errors, three of which are associated with the VACS and one with the conning officer. They are described as follows:

- **Type 1 error: SRS uses the wrong word**

In this instance a misinterpretation of the verbal input was made by the SRS and an incorrect word was substituted for the appropriate word.

- **Type 2 error: SRS adds a word not spoken**

This error occurs when the SRS believes a word is uttered when it was not. Some instances where this error type may occur include inadvertent contact with the microphone causing a crackle noise, clearing of the throat, or any other superfluous background noise detected by the microphone and transmitted to the SRS.

- **Type 3 error: SRS does not acknowledge a spoken word**

In this case, the SRS fails to receive the incoming acoustic signal. Some extraneous causes of this error include a microphone failure, overpowering background noise, a very soft spoken or extremely brief/fast verbal signal.

- **Type 4 error: A nonstandard command is used**

It is possible for the conning officer, if not properly trained, to use an incorrect command format thereby representing improper syntax for the SRS to interpret. This occurs because the conning vocabulary stored in the SRS memory contains only the words and phrases of standard commands. As a result the SRS will be unable to correctly identify a word not contained in the restricted vocabulary list.

The proper use of standard commands is paramount on the bridge of any naval vessel and only through extensive training does a conning officer become proficient. Because Type Four errors reflect insufficient conning officer training, these errors are not associated with measurements of SRS effectiveness. The remaining three types of errors are suitable metrics for SRS, and are aggregated for this analysis. The reason for combining the errors rather than examining each type independently is to reflect the

overall system performance and so that data from this study can be compared to the earlier research in which the error types were also combined.

B. DATA ANALYSIS

Before performing statistical analysis of the experiment outcomes it is necessary to ensure that the experiment meets all the prerequisites of sound design. As mentioned in Chapter III, randomizing the sequence in which the scenarios were presented and the assignment of test subjects to time slots removes the chance that a certain order of events could affect the experiment outcome. In other words, randomization ensures that chance governs the results and not any characteristic of the experimental procedure or the judgment of the experimenter. [Ref. 44] Having randomized, the next step is to determine if normality existed. The result of this analysis is depicted in a standard normal quantile plot, using statistical software package S-Plus. [Figure 2]

> The adequacy of a normal model for describing a distribution of data is best assed by a normal quantile plot. A pattern on such a plot that deviates substantially from a straight line indicates that the data are not normal. [Ref. 45]

The horizontal axis of this plot is numbered from -2 to 2. The zero point represents the median data point. On either side of this median value are the next higher or lower values. In this case, the standard normal quantile plot shows a relatively normal distribution of SRS errors. A normal distribution is one in which the data points begin in the lower left corner of the graph and follow an imaginary line to the upper right corner of the graph. Normal distributions approximate the outcomes of chance. This is an important factor because any further analysis of variance (ANOVA) or sample mean testing requires a normally distributed variable. [Ref. 46] These statistical inference procedures rely on normal distributions to calculate the mean and standard deviation without the influence of any outliers or other non-standard results.

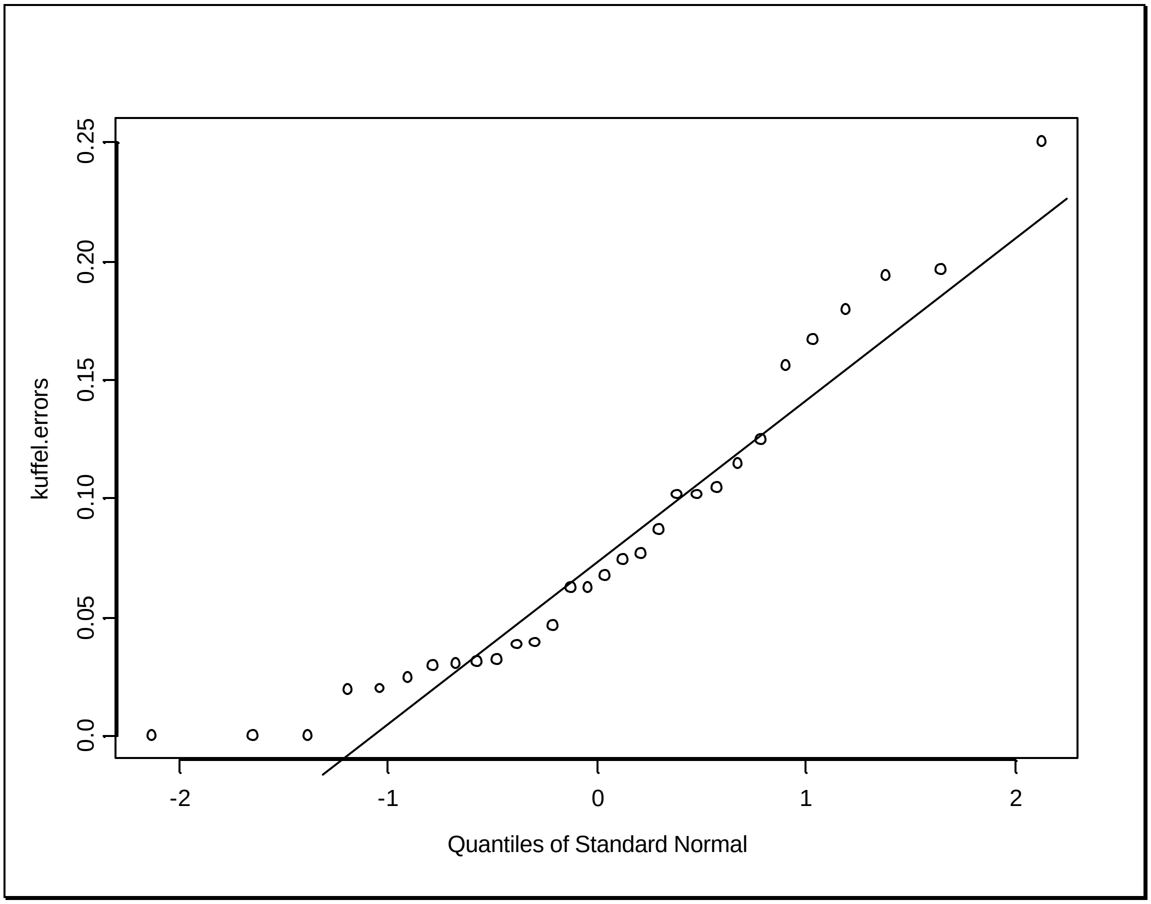

Figure 2. Standard Normal Quantile Plot

ANOVA methodology examines explained and unexplained variations in performance measures to determine the significance of a model. In other words, we are looking to see if the variation between the experimental results and the normally expected results differ. The distance between the data points and their mean value is the variation. The measure of variation is the distance between observation and expectation, tallied by summing the squared differences.

Dividing the sums of squares by the appropriate degrees of freedom yields an estimation of mean square differences. The ratio of the mean squares is an F-statistic that shows the average amount of explained variation as compared to the average amount of unexplained variation. The larger the F-value, the more explained variation and the less unexplained. If there is no explanatory relationship then the ratio of explained variation

to unexplained variation will be small. This supports the null hypothesis which assumes no significant explanation of observed performance. [Ref. 47]

After computing the F-statistic, based on the observed data, and comparing this value to the known F distribution, analysis yields a P-value. This P-value is referred to as "Pr(F)" in the analysis charts that follow. The P-value is the probability of observing the results seen during the experiment given that the null hypothesis is true. The null hypothesis states that introduction of an explanatory variable will not have an effect on the performance responses of the study. [Ref. 48] As discussed above, the null hypothesis is that there is no difference in SRS performance among groups based on vocabulary size, experience level or training. Armed with this knowledge we can apply ANOVA to each of the expectations outlined in Chapter III specifying whether evidence supports or refutes the null hypothesis.

C. EXPECTATION AND DATA COMPARISONS

1. Expectation #1

The first expectation for this study is that SRS performance with smaller vocabularies is more accurate (fewer errors) than SRS performance with large vocabularies. The analysis of this expectation required that the results from the earlier SRS study using a large vocabulary be compared to the data obtained in this restricted vocabulary study. The results of this analysis are in Table 4 and clearly indicate that there is no significant difference in performance among users of the restricted and the large vocabulary. The likelihood of seeing these outcomes if there were no difference in SRS performance based on size of vocabulary is relatively high.

E1	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
Exp vocab	1	0.00130975	0.001309746	0.6811449	0.4147828
Residuals	35	0.06730009	0.001922860		

Table 4. ANOVA for Expectation 1 (Vocabulary Size)

One possible reason for the absence of a significant difference between vocabulary sizes has to do with the test procedure. According to the DNSV6.0 user's manual, the fastest way for the software to "learn" is for users to make corrections to errors as they occur. The testing procedure used in this experiment did not include the use of the DNSV6.0 correction capabilities. Test subjects made no corrections during the experiment, potentially lengthening the learning curve for the software. In fact, a review of the raw data shows several instances where the same test subject produced errors on the same phrase multiple times. Perhaps using the correction function after the original error averts follow-on errors.

Taking the level of accuracy into consideration further explains these results. With either a restricted vocabulary or a large vocabulary DNSV6.0 is ninety to ninety-nine percent accurate in most trials. The reduced vocabulary trials make it easier and possibly faster for the software to match words to spoken language, but do not necessarily reduce errors caused by poor pronunciation, background noise, and other non-SRS related factors. A small percentage of non-SRS related errors occur in each trial. These are not eliminated by reducing the size of the vocabulary. The exact number of non-SRS related errors varies from subject to subject and therefore is not accounted for in this experiment. However, while the reduced vocabulary SRS did not significantly reduce the number of errors, there are other benefits to using an SRS with a small vocabulary. The reduced processing time associated with small vocabulary SRS makes the software more efficient and responsive. This potential benefit alone makes smaller vocabulary SRS more desirable for highly dynamic applications.

Finally, as reported in Chapter I, SRS uses statistical language models to predict the likelihood of a word occurring in a sentence. This experiment however, did not use normal sentence structure and grammar. It used standard naval commands which do not conform to the rules of the statistical language model. The statistical model therefore lost some of its predictive power.

2. Expectation #2

Expectation number two states that the experience level and/or gender of the SRS user has no impact on the performance of SRS. While it was always the aim of this study to examine the role of experience level, examining the issue of gender was not part of the original design. The comment of a simulator operator at Surface Warfare Officer's School led to the incorporation of the gender issue. One of the SWOS instructors stated that the system seemed to make more errors with female operators than it did with male. [Ref. 49] Independent and combined analysis of these variables was completed to ensure there are no confounding effects.

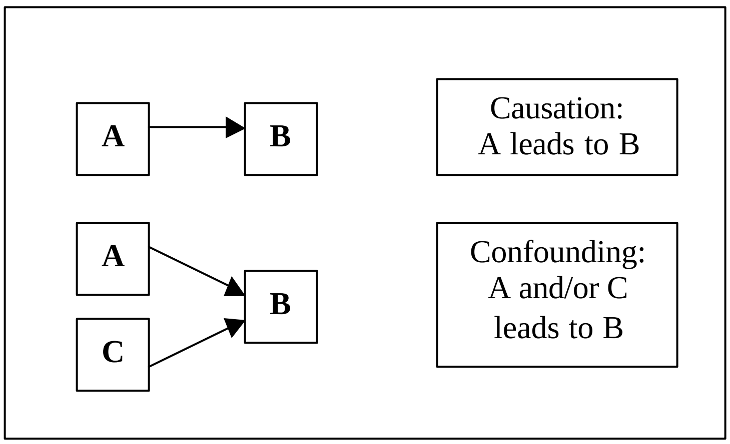

Figure 3. Confounding Variables

Two or more variables are confounded when their effects are mixed together. [Ref. 50] Tables 5 through 7 show the analysis of experience level, gender and then gender and experience, respectively.

E2	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
Experience	1	0.0028421	0.002842133	0.6373282	0.4313994
Residuals	28	0.1248646	0.004459450		

Table 5. ANOVA for Expectation 2 (Experience Level)

E2	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
Gender	1	0.0144058	0.01440583	3.637775	0.06678888
Residuals	28	0.1108818	0.00396007		

Table 6. ANOVA for Expectation 2 (Gender)

E2	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
Gender	1	0.0144058	0.01440583	3.643163	0.0669833
Experience	1	0.0041182	0.00411819	1.041471	0.3165372
Residuals	27	0.1067636	0.00395421		

Table 7. ANOVA for Expectation 2 (Experience and Gender)

As seen above, there is no significant difference between test subjects based solely on experience level. A P-value of .431 fails to refute the null hypothesis that the experience level of the conning officer does not impact SRS performance. Gender however, is a significant factor regardless of the experience level. (P-value of .066) It should be noted though, that the sample size is insufficient to make serious SRS generalizations. The sample size determines the margin of error and with only two female test subjects our margin of error is very high. [Ref. 51] A larger female sample size was not obtained as stated above, because the original focus of this study did not include the issue of gender.

The results of the gender and experience level analysis (Figure 3) show a general increase in the error rate as one moves from experienced males to inexperienced males to females. However, because there were no high experience level female test subjects the possibility of confounding variables exists. Because all of the female test subjects were inexperienced, it is unclear if the observed effects are due solely to their gender or a combination of low experience and gender. Further research in this area is necessary to separate the two variables. Another noticeable trend in Figure 3 is that there is a greater

degree of variability among the female test subjects. There is no apparent cause for this increased data spread and again serves to show that additional research with female SRS users is warranted.

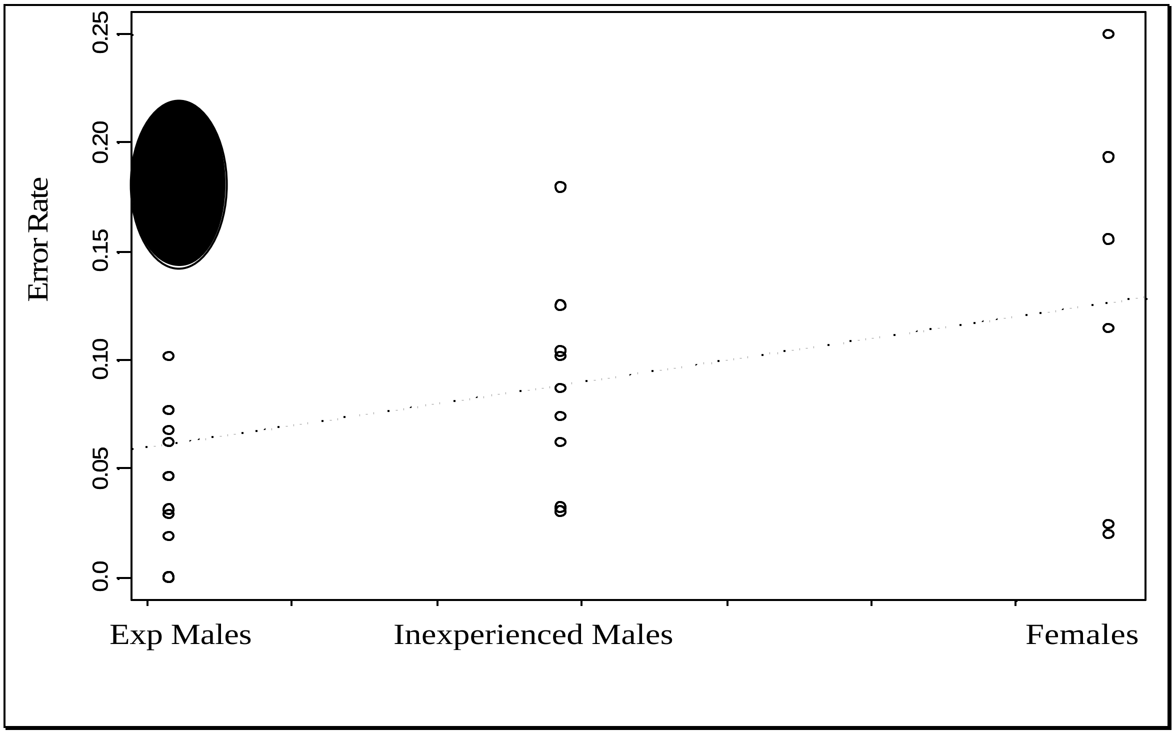

Figure 4. Experience Level And Gender Plot

The dotted line shown in Figure 4 represents the median values in each category. The two trials with the highest error rate in the experienced male category (circled) were caused by a single test subject with a heavy New England accent. These two outliers increase the mean, but not the median. The median, a resistant measure against outliers, shows an increase as experience level decreases. This is not significant however because the means and errors used in the F-test are not resistant to outliers. Using the median values solidifies the proposition that gender impacts SRS performance.

Without taking gender in to account however, there is no significant difference between test subjects based solely on experience level. There may appear to be a

difference in Figure 4, but using the mean values, it is not significant. This confirms our expectations and demonstrates that SRS performance has little to do with the experience level of the conning officer.

3. Expectation #3

Expectation number three suggested that SRS training has no impact on error rates**.** Figure 5 depicts a side by side box plot and shows a significant correlation between error rates and training supporting our expectation. The full details of the additional SRS training were presented in Chapter III.

Figure 5. Training vs. No Training Error Rates

According to this analysis, with a P-value of .05, SRS capable of individual user training will produce fewer errors. This is an important design characteristic to consider for future SRS implementation. Current Navy training simulators with SRS technology do not use individual user training. [Ref. 52] It would seem however that any SRS system designed for ship control purposes should incorporate a user training feature.

Another noticeable difference between the two sets of data is the spread of the results. The small white boxes that surround the median error rate represent the interquantile range in which fifty percent of the data falls. Notice that the "training" white box is the smaller of the two and that there is no overlap in the area covered by the boxes. The "no training" group consists of only nine trials while the "training" group had

21 trials. The data are much more tightly grouped in the "training" trials, despite the larger number of trials. This indicates that additional training with SRS eliminates some of the variation and produces a more accurate and well-defined result.

As encouraging as these results are, it must be noted that the method of training used in this study is not thorough enough. The training is limited to only two words; "starboard" and "rudder". Test subjects repeated each of these words several times until the software established firm models for each word. These words were selected due to their high rate of error observed in pre-trial exercises. However, to truly assess the value of system training, the training should include most if not all of the words and phrases in the restricted conning vocabulary. A more comprehensive training method may reduce error rates even further.

V. CONCLUSION

A. SUMMARY

The U.S. Navy of today emphasizes cost cutting through reductions in manpower numbers. One of the key components enabling this reduction of the workforce is the substitution of technology for watchstanders. This SRS study shows the practicality of using commercial-off-the-shelf speech recognition software for ship control purposes thereby demonstrating the prospect of eliminating bridge watchstanders. SRS is already used in the military training environment and could be adapted for operational use as well. Despite the fact that the technical feasibility of SRS implementation is very high, a number of questionable psychological barriers remain and may only be overcome through the proven reliable usage of SRS over time. Previous experimentation with SRS led to the identification of several areas that required follow-on research. This study, through the use of a controlled experiment using COTS SRS addresses many of those outstanding areas. The results of this study show that:

- The experience level of a conning officer has no impact on SRS performance; however, in this experiment, limited number of trials indicate that gender may make a difference. Female participants experienced more SRS errors than did their male counterparts.
- SRS with restricted vocabulary performs no better than SRS with large vocabularies.
- Following the user profile establishment, individual user training on two specific words reduces error rates significantly.

B. LIMITATIONS OF STUDY

Some of the limitations of this study are found by examining the testing environment. While the use of the MSI simulator facility in Newport Rhode Island was conducive to experimentation, the simulator does not capture all of the nuances of an actual shipboard environment. Background noises such as additional watch stander conversations, ships and tug boats whistles, wind noise were not examined and may impact the performance of a COTS SRS product. Additionally, the use of a wireless microphone in the simulator was made possible due to the lack of competing signals. Onboard ship, the radio frequency environment could cause signal conflicts.

A second limitation is the small number of test subjects. A larger pool of test subjects would increase the power of obtained results. A major shortcoming was that only two of the test subjects were female.

C. PROPOSED FOLLOW-ON RESEARCH

Due to these limitations as well as new insight gained during this study, there is a need for additional research in the area of COTS SRS used for ship control purposes. The paragraphs below propose several areas for follow-on research.

- Use a large pool of both high and low experience female test subjects to determine the impact that gender has upon SRS performance.
- Conduct tests underway aboard actual naval vessels to determine the impact of shipboard background noise on SRS.
- Match high and low stress in experiments to gain insight into the role that excitement and changes in voice pitch play in SRS accuracy.
- Consider SRS processing speed as a measure of performance and research the processing time of a large vocabulary SRS versus the processing time of a restricted vocabulary SRS.
- Study SRS user training to determine its benefit, specifically analyzing whether training should be conducted on all words in a restricted vocabulary SRS.

Based on the results of this study, with further testing and development COTS SRS is a viable alternative to reduce shipboard manning if it incorporates individual user training, redundancies, and safe-guards as discussed in Chapters II and III. Its initial use could be limited to open ocean transits until the Navy gains confidence in eliminating the helmsman and lee helmsman watchstanders. Shore-based SRS ship handling simulators like the ones currently in use continue to expose and train new "ship drivers" to the intricacies of SRS use. These measures can help to ensure a smooth transition to SRS based ship control. The Chief of Naval Operations guidance and the Naval Transformation Roadmap both endorse inserting technology to develop manpower-saving capabilities. Speech Recognition Software is precisely the type of technology that can fulfill this requirement.

APPENDIX A. STANDARD COMMANDS

Standard commands will vary depending on the type of ship. Listed below is the format for the most common standard commands used by naval surface vessels.

Engine orders:

WHICH ENGINE[1], *DIRECTION*[2], *AMOUNT*[3]

WHICH ENGINE[1] stop.

1. Starboard engine / Port engine / All engines
2. Ahead / Back
3. 1/3 / 2/3 / standard / full / flank / or by pitch (i.e. "20% pitch")

Steering orders:

DIRECTION[1], *AMOUNT*[2], steady on *COURSE*[3,4] - (Used for course changes greater than 10 degrees)

Come *DIRECTION*[1], steer course *COURSE*[3] – (Used for course changes less than 10 degrees)

Hard *DIRECTION*[1] rudder, steady on *COURSE*[3,4] – (Used for extremis steering)

1. Right / Left
2. Standard rudder / full rudder / or number of degrees (i.e. "10 degrees rudder")
3. Any heading between 000 and 359.
4. A steady on course is optional.

Additional standard commands used for steering:

Rudder amidships
Steady as she goes
Meet her
Mind your helm
Shift your rudder
EASE or INCREASE your rudder to *DIRECTION*[1] , *AMOUNT*[2]

1. Right / Left
2. Standard rudder / full rudder / or number of degrees (i.e. "10 degrees rudder")

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. MSI/NPS TEST DOCUMENT

MARINESAFETY INTERNATIONAL INTER-OFFICE MEMORANDUM

10/24/03

To: Distribution
From: Fred Bronaugh, CAPT USN (Ret.)
Subject: US Naval Post Graduate School Voice Recognition Experiment (change 2)

1. From Monday the 27th of October to Wednesday the 29th we will be hosting a Voice Recognition experiment for the NPGS. Lt Rob Kuffel will be test director and will be using very experienced mariners (us) and experienced mariners (SWOS instructors) in the experiments.

2. The BWS will be the experiment site and three different docking scenarios in NCH will be used. Expect the three will be (A) moor and U/W from 2S, (B) moor and U/W from 3P and (C) Moor and U/W from 7S. All runs will be no current, no wind and will start about two ship lengths from the pier. The CG-47 class will be the own ship.

3. Schedule of events and tasking:

Monday 27OCT	Test Subject	Sequence
0800-0900: Set-up		
0900-1100: TEST SUBJECT A	LT Reichenau	[A,B,C]
1100-1200: Lunch		
1200-1400: TEST SUBJECT B	LT Mullins	[B,A,C]
1400-1600: TEST SUBJECT C	Dan Liuzzi	[C,B,A]
1600-1700: Flex time		
Tuesday 28OCT		
0800-1000: TEST SUBJECT D	Bud Weeks	[C,A,B]
1000-1200: TEST SUBJECT E	Dave Kane	[A,B,C]
1200-1300: Lunch		
1300-1500: TEST SUBJECT F	Ed Lynch	[B,A,C]
1500-1700: TEST SUBJECT G	LT Rickwalt	[C,B,A]
Wednesday 29OCT		
0800-1000: TEST SUBJECT H	Fred Bronaugh	[A,C,B]
1000-1200: TEST SUBJECT I	LT Balcirak (female)	[B,C,A]
1200-1300: Lunch		
1300-1500: TEST SUBJECT J	LTjg Krug (female)	[A,B,C]
1500-1700: Flex time/Wrap-up		

4. The operator will maintain control of rudder and engines, commands will be relayed by hand/headset. The objective is to evaluate the effectiveness and reliability of the software not to evaluate shiphandling skill. Setup will be the responsibility of Lt Kuffel, Calvin you should be ready to provide assistance.

Thanks
Fred

Distribution: Ed. Bud. Dan L. Dave. Pete. George K. Tom. Jim and Calvin

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. TEST SUBJECT BRIEF

Thank you for agreeing to participate in this study. The purpose of this study is to determine the reliability of commercial off the shelf speech recognition software when used for ship control purposes. You will be asked to conn a simulated CG in or out of port while wearing a wireless microphone. It is important that you remember that your ship driving ability is NOT being tested. Try to remain calm throughout your scenario and speak in a loud and clear voice. Try to avoid contact with the microphone and external conversations. If you must say something other than an engine or rudder order you may switch off the microphone temporarily. When you turn it back on however, be sure to pause before giving an order. Your verbal commands will be transmitted to a laptop computer that will convert them into text. The entire experiment should take about 2 hours.

The first step will be to set-up a user profile on the computer. [SET UP PROFILE]

The format for standard commands that you should use for this experiment is as follows:

Engine orders:

WHICH ENGINE[1], *DIRECTION*[2], *AMOUNT*[3]

WHICH ENGINE[1] stop.

1. Starboard engine / Port engine / All engines
2. Ahead / Back
3. 1/3 / 2/3 / standard / full / flank / or by pitch (i.e. "20% pitch")

Steering orders:

DIRECTION[1], *AMOUNT*[2], steady on *COURSE*[3,4] **-** (Used for course changes greater than 10 degrees)

Come *DIRECTION*[1], steer course *COURSE*[3] – (Used for course changes less than 10 degrees)

Hard *DIRECTION*[1] rudder, steady on *COURSE*[3,4] – (Used for extremis steering)

1. Right / Left
2. Standard rudder / full rudder / or number of degrees (i.e. "10 degrees rudder")
3. Any heading between 000 and 359.
4. A steady on course is optional.

All other standard commands remain unchanged.

Rudder Amidships
Ease or increase your rudder to …
Steady as she goes
Etc.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D: SRS SPECIALIZED VOCABULARY

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
All
All engines ahead
All engines back
Amidships
Rudder amidships
And
As
Back
Course
Come
Come right steer course
Come left steer course
Degrees
Ease
Ease your rudder to left
Ease your rudder to right
Engine
Engines
Flank
For
Full
Goes
Hard
Hard right rudder
Hard left rudder
Helmsman
Increase
Increase your rudder to right
Increase your rudder to left
Indicate
Knots
Left
Left full rudder
Left standard rudder
One third
Percent
Percent pitch
Pitch
Port
Port engine ahead
Port engine back
Right
Right full rudder
Right standard rudder
Rudder
Shift
She
Standard
Starboard
Starboard engine ahead
Starboard engine back
Steady
Steady as she goes
Steer
Stop
To
Turns
Two thirds
You
Your

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. TEST DATA SPREAD SHEET

All Error Types

Summary Data

trial	# errors	# orders	P[error]	subj	exp level	gender	scenario	sequence
1	4	48	0.083333	I	low	M	A	1
2	4	46	0.086957	I	low	M	B	2
3	4	54	0.074074	I	low	M	C	3
4	5	40	0.125	II	low	M	B	1
5	5	48	0.104167	II	low	M	A	2
6	6	59	0.101695	II	low	M	C	3
7	5	48	0.104167	III	high	M	C	1
8	3	32	0.09375	III	high	M	B	2
9	3	28	0.107143	III	high	M	A	3
10	13	72	0.180556	IV	high	M	C	1
11	7	59	0.118644	IV	high	M	A	2
12	11	51	0.215686	IV	high	M	B	3
13	6	39	0.153846	V	high	M	A	1
14	1	33	0.030303	V	high	M	B	2
15	4	43	0.093023	V	high	M	C	3
16	1	48	0.020833	VI	high	M	B	1
17	2	51	0.039216	VI	high	M	A	2
18	3	52	0.057692	VI	high	M	C	3
19	1	52	0.019231	VII	low	M	C	1
20	1	34	0.029412	VII	low	M	B	2
21	0	32	0	VII	low	M	A	3
22	3	33	0.090909	VIII	high	M	A	1
23	11	39	0.282051	VIII	high	M	C	2
24	3	31	0.096774	VIII	high	M	B	3
25	7	31	0.225806	IX	low	F	B	1
26	8	45	0.177778	IX	low	F	C	2
27	3	12	0.25	IX	low	F	A	3
28	7	61	0.114754	X	low	F	A	1
29	1	41	0.02439	X	low	F	B	2
30	2	50	0.04	X	low	F	C	3

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

1. Liewer, Steve. “Mobile Bay Among First to Participate in Reduced-crew Project.” *Stars and Stripes*. August 3, 2001.

2. Wilson, Shawn C. “Voice Recognition Systems: Assessment of Implementation Aboard U.S. Naval Ships.” Master’s Thesis. Naval Postgraduate School, Monterey, CA, March 2003. p. 5

3. Zue, V., Cole, R. and Ward, W., Survey of the State of the Art in Human Language Technology, 1996, [cslu.cse.ogi.edu/HLTsurvey/ch1node4.html#parameters], October 4, 2003.

4. Markowitz, Judith A. Using Speech Recognition. Prentice Hall, NJ, 1996. p. 231.

5. Markowitz, Judith A., p. 245.

6. Tamez, D.J. “Using Commercial-Off-the-Shelf Speech Recognition Software for Conning U.S. Warships.” Master’s Thesis. Naval Postgraduate School, Monterey, CA, June 2003. p. 18

7. West, Timothy J. “Implementation and Evaluation of Commercial Off-the-Shelf (COTS) Voice Recognition Software as an Input Device in a Windows-Type Environment.” Master’s Thesis, Naval Postgraduate School, Monterey, CA, March 1996. p. 12.

8. West, Timothy J., p. 13.

9. Tamez, D.J., p. 19.

10. West, Timothy J., p. 2.

11. United Cerebral Palsy of New York City, Environmental Remote Control Units, [www.ucpnyc.org/info/assist/environment.cfm], October 14, 2003.

12. Wilson, Shawn C., p.14.

13. Lawlor, Maryann. “Topside Training Submerges Students In Virtual Reality.” *SIGNAL Magazine*, July 2001.

14. Email message between John Mentzer, Lieutenant USN and the author, 27 October 2003.

15. Ingall’s Shipbuilding, Inc., DTN Risk Reduction Test DTIB-509 Revision, September 28, 2001.

16. Tamez, D.J. “Using Commercial-Off-the-Shelf Speech Recognition Software for Conning U.S. Warships.” Master’s Thesis. Naval Postgraduate School, Monterey, CA, June 2003.

17. Tamez, D.J., p. 41.

18. Tamez, D.J., p. 52.

19. Stavridis, James. U.S. Navy Watch Officer’s Guide. United States Naval Institute, 14th edition. December 1999. p. 123.

20. Wilson, Shawn C., p.10.

21. Elbert, George. “Clinton is a Shameless Commander in Chief” *Honolulu Star Bulletin*, May 31, 1999.

22. Binner, Tom. “ORM Corner: A Bad Year We Could Have Prevented,” *Fathom Magazine,* January-March 2001.

23. England, G., Clark, V. and Jones J.L., Naval Transformation Roadmap: Power and Access From the Sea, p. 5 [www.seabee.navy.mil/nmcb1/Batallion%20News/archives/Naval%20Trans%20Roadmap%20_Final_2.pdf] as of September 23, 2003.

24. Interview between Richard West, Rear Admiral, USN, navigator of the Navy, and *Fathom Magazine,* April-June 2002.

25. Clark, Vern. “Accelerating Our Advantages”, *CNO 2004 Guidance.* January 2004.

26. Brown, David. “Smaller Smarter Fleet; CNO Vern Clark’s plan for a leaner Navy could affect every sailor.” *Navy Times*, 19 January 2004. p.14

27. Brown, David., p.15

28. Wilson, Shawn C., p.15

29. Wilson, Shawn C., p. 17.

30. Bost, J. Robert, James Mellis and Phillip Dent. “Is the Navy Serious About Reducing Manning on its Ships?” 1999. [www.manningaffordability.com/S&tweb/Index_main.htm] 12 October 2003.

31. Bost, J. Robert, James Mellis and Phillip Dent. “Is the Navy Serious About Reducing Manning on its Ships?” 1999. [www.manningaffordability.com/S&tweb] 13 October 2003.

32. Wilson, Shawn C., p. 29.

33. Wilson, Shawn C., p. 31.

34. Wilson, Shawn C., p. 27.

35. Wilson, Shawn C., p. 24.

36. Newman, Dan. *Dragon NaturallySpeaking6 User's Guide,* ScanSoft, Inc., May 2002. p. 221.

37. Tamez, D.J., p. 33.

38. Marine Safety International, [www.marinesafety.com/sections/site/aboutus.html] as of October 13, 2003.

39. Tamez, D.J., p. 31.

40. Tamez, D.J., p. 33.

41. Environmental Protection Medical Unit Newport, Rhode Island noise level survey conducted 26 October 2003.

42. Voice Recognition Software, Consumer Search: Reviewing the Reviewers, The Wall Street Journal Online, [wsj.consumersearch.com/computers/voice_recognition_software/fullstory.html] as of September 28, 2003.

43. Newman, Dan., p. 146.

44. Moore Davis S. and McCabe, George P. Introduction to the Practice of Statistics, Third Ed. W.H. Freeman and Company. 1999. p. 241.

45. Moore Davis S. and McCabe, George P., p. 85.

46. Moore Davis S. and McCabe, George P., p. 803.

47. Moore Davis S. and McCabe, George P., p. 271.

48. Moore Davis S. and McCabe, George P., p. 753.

49. Conversation between John Mentzer, Lieutenant USN and the author, 27 October 2003.

50. Moore Davis S. and McCabe, George P., p. 209.

51. Moore Davis S. and McCabe, George P., p. 443

52. Email message between John Mentzer, Lieutenant USN and the author, 27 October 2003.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Dr. Dan Boger
   Chairman, Department of Information Sciences
   Naval Postgraduate School
   Monterey, California

4. Dr. Monique P. Fargues
   ECE Dept. Code EC/Fa
   Naval Postgraduate School
   Monterey, California

5. LCDR Russell Gottfried, USN
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

6. LT Robert Kuffel, USN
   SWOS Department Head School
   Newport, Rhode Island

7. CAPT J. Kline, USN
   Chairman Warfare Innovation
   Naval Postgraduate School
   Monterey, California

8. CAPT Fred Boronough, USN (Ret.)
   Center Director
   Marine Safety International
   Newport, Rhode Island

9. Mr. T. Derrico
   ScanSoft, Inc.
   Worldwide Headquarters
   Peabody, Maryland

10. Dr. Judith Markowitz
    President, J. Markowitz Consultants
    Associate Editor, Speech Technology Magazine
    Chicago, Illinois

11. RADM David Ramsey, USN (Ret.)
    Center Director
    Marine Safety International
    San Diego, California